DOCUMENT RESUME

ED 046 666 RE 003 286

AUTHOR	Follman, John; And Others
TITLE	Psychometric Analysis of Critical Reading and Critical Thinking Tests--Twelfth Grade.
PUB DATE	Dec 70
NOTE	9p.; Paper presented at the National Reading Conference, St. Petersburg, Fla., Dec. 3-5, 1970
AVAILABLE FROM	Twentieth Yearbook of the National Reading Conference, Inc., Marquette University, 1217 W. Wisconsin Ave., Milwaukee, Wis. 53233
EDRS PRICE	EDRS Price MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	*Critical Reading, *Critical Thinking, Factor Analysis, Grade 12, *Psychometrics, Reading Research, *Reading Tests, *Test Reliability, Test Validity, Thought Processes

ABSTRACT

Psychometric characteristics and factor structure of the Reading Comprehension Test (CR) and the Test of Critical Thinking, Form G (CT) were studied in an effort to determine basic dimensions of critical reading and critical thinking and to discover the relationship between these two. The two tests were administered to 57 Florida high-school seniors. Total test score reliability estimates were .88 and .92 for CR and .86 and .87 for CT. Factor analysis indicated that the tests measured separate skills. It was concluded that (1) both tests are psychometrically sound, (2) CR represents a relatively homogeneous underlying variable while CT represents a number of different variables, (3) critical reading is a thinking ability involving judgment of materials and critical thinking is less clearly defined and includes judgments of how conclusions are reached, and (4) critical reading and critical thinking as represented by the tests studied overlap only moderately. References are included. (Author/MS)

EDO 46666

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

National Reading Conference

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Psychometric Analysis of Critical Reading and Critical Thinking Tests - Twelfth Grade

John Follman A. J. Lowe Wade Burley Johnny Follman

University of South Florida

Introduction

This study is the second in a series of three statistical analyses of critical reading test score data of twelfth grade students. The twelfth grade data represents half of an overall empirical examination of the definition of critical reading and its relationship with critical thinking, reading, intelligence, and achievement test scores. The other half of the overall empirical examination of the definition of critical reading is a parallel series of three analyses of scores of similar tests from fifth grade pupils.

RE003 286

In the first analysis of the twelfth grade test scores Follman, Lowe, and Wiley (2) found that critical reading overlaps substantially with reading, thinking, and language activities, particularly vocabulary, and also with critical thinking.

The objective of this, the second study in the twelfth grade series, was to investigate in depth the psychometric characteristics of the critical reading test, and the critical thinking test so that precise inferences could be made about the definition of critical reading and its relationship with critical thinking.

The tests analyzed were:

Reading Comprehension Test (CR) (Martin, 4) total
- Main Points (MAIN PTS) subtest
- Specific Facts (SPEC FACTS) subtest
- Cause and Effect (CAUSE EFFECT) subtest
- Inference (INFERENCE) subtest
- Vocabulary (VOCABULARY) subtest

Test of Critical Thinking Form G (CT) (ACE, 1) total
- Pertinent Information (PERT INFO) subtest
- Valid Inferences 1 (VAL INF 1) subtest
- Valid Inferences 2 (VAL INF 2) subtest
- Relevant Generalizations (REL GENS) subtest
- Recognition of Assumptions (RECOG ASSUMP) subtest
- Valid Inferences 3 (VAL INF 3) subtest
- Valid Inferences 4 (VAL INF 4) subtest
- Hypothesis Verification 1 (HYP VER 1) subtest
- Hypothesis Verification 2 (HYP VER 2) subtest

The Reading Comprehension Test was the critical reading test.

Procedure

The subjects (Ss), twelfth grade students from Robinson High School, Hillsborough County, Florida, were tested in the fall of 1969. The Ss were selected to represent typical twelfth grade Robinson High students. Mean IQ was 100.66.

Item difficulty and discrimination indices were determined.

Split-half odd even correlation and Kuder-Richardson 20 reliability estimates were determined for subtest and total test scores.

The basic dimensions of CR, CT, and CR and CT combined, the critical reading test, the critical thinking test, and the critical reading and critical thinking tests combined, were investigated through inter-item phi coefficients, principal components factor analysis and rotation of factors with eigenvalues in excess of one. N was 57 for all analyses.

Results

Mean item difficulty and discrimination indices were .42 for both CR and CT. A few items for each measure did not discriminate and should either be eliminated or refined.

Table 1 indicates subtest item groupings, medians, means, standard deviations, odd even split half and Kuder-Richardson 20 reliability estimates for subtest scores and total test scores for CR and CT.

Table 1

Medians, Means, Standard Deviations, Odd Even and Kuder-Richardson Reliability Estimates

		Items	Mdn	$\bar{X}$	SD	OE	KR
CR							
	MAIN PTS	1-5	3	3.47	1.45	.74	.56
	SPEC FACTS	6-20	13	12.82	2.51	.71	.75
	CAUSE EFFECT	21-25	3	3.39	1.41	.49	.53
	INFERENCE	26-35	6	6.21	2.17	.63	.57
	VOCABULARY	36-50	11	11.44	2.93	.76	.76
	TOTAL	1-50	38	37.30	8.90	.88	.92
CT							
	PERT INFO	1-6	4	4.42	1.77	.74	.66
	VAL INF 1	7-10	1	1.32	1.04	.43	.20
	VAL INF 2	11-16	1	2.75	1.63	.66	.53
	REL GENS	17-21	1	2.53	1.32	.30	.34
	RECOG ASSUMP	22-31	5	6.18	1.81	.43	.40
	VAL INF 3	32-34	1	1.23	1.13	.50	.46
	VAL INF 4	35-37	1	1.04	.98	.52	.36
	HYP VER 1	38-46	2	4.05	2.21	.40	.63
	HYP VER 2	47-52	1	2.53	2.10	.87	.69
	TOTAL	1-52	27	27.86	8.64	.86	.87

Total test score reliability estimates were high for both CR and CT, and subtest score reliability estimates with one exception were above .29 with most considerably higher.

Inter-item phi coefficients for CR ranged from .65 to -.50 with many non-significant. Correlations of .26 and .34 were significant at the .05 and .01 levels respectively. No correlation matrices or factor loading tables are presented because of space limitations. However, these tables are available upon request.

Factor analysis of the 50 x 50 inter-item phi matrix for CR indicated several group factors. Considering loadings of .30 or greater the first factor consisted of 22 items and accounted for 20% of the total test variance. Subsequent factors consisted of successively fewer items and accounted for successively smaller amounts of total test variance. Eigenvalues were successively

9.9, 5.1, 3.7, etc. Apart from the first factor, which had loadings from items of all five subtests, items tended to load in sets partly consistent with the a priori subtest item groupings in that there were several groups of items from their respective a priori subtest groupings. They were also inconsistent in that these sets of items often represented two or more subtests and loaded on the same respective factors.

Rotation demonstrated remarkable consistency of both items and item strengths for the largest group factor. Twenty items loaded on both factors across rotation. These items apparently measure the same underlying construct. However, the items generally did not cluster in the a priori subtest groupings categorized by Martin (4). Other factors generally were not consistent across rotation nor consistent with the a priori subtest item groupings although some items grouped together consistent with the a priori groupings. Point biserial correlations between each item and total test score ranged from .102 to .696 with a median of .467. Most of the items with high correlations were the items that loaded on the largest group factor, additional evidence that many CR items measure the same underlying construct.

Inter-item phi coefficients for CT were generally lower than those of CR and most were between ±.35 with many lower non-significant ones.

Factor analysis of the unrotated and rotated factor loadings of the 52 x 52 inter-item phi matrix for CT indicated mostly factor loadings of moderate or low strengths. Eigenvalues for the respective factors were successively 5.1, 3.8, 3.0, etc. Factor analysis indicated more small group factors consisting of both fewer items and particularly weaker loadings than the CR factors.

The smaller loadings by fewer items per factor were accentuated by the rotation as it produced high loading individual items many of which appeared with two or three other relatively weak item loadings. Apparently most CT items represent different underlying variables which generally correlate low but which correlate higher with the overall critical thinking construct (or

whatever CT measures) as reflected by the high total test score reliability. Point bi-serials ranged from -.109 to .612 with a median of .379. These correlations were generally considerably lower than those for CR, and are also additional evidence that the CT items measure a number of different variables. There were relatively few item groupings consistent with the test makers' (ACE, 1) a priori subtest categorizations.

In order to determine the relationship between critical reading and critical thinking, an analysis of the 102 x 102 item matrix of both tests was conducted. The CR items generally intercorrelated low or moderately as did the CT items but the CR and CT items generally correlated lower than the items within either CR or CT only.

Factor analysis indicated interesting variance structure. Small group factors appeared accounting respectively for 12%, 7%, and successively smaller amounts of variance. The first group factor, accounting for 12% of the variance, consisted of low or moderately sized loadings from 10 CT items and low, moderate, and high loadings from 27 CR items. Subsequent group factors had similar but fewer item compositions from both tests. The lower strengths of the CT items vis a vis the CR items is noteworthy, apparently reflecting somewhat the idiosyncratic nature of the CT items as well as the greater common variance of the CR items.

Rotation in the 102 x 102 analysis dramatically revealed the extent of the split between CR and CT variance. Rotation produced a number of smaller group factors which were with infrequent exceptions specific to CR items or CT items, not items from both CR and CT on the same factors. This helps account for the fact that no general, or large group, or even meaningful small group factor appeared in terms of amount of variance accounted for. This apparently means that the CR items represent CR test specific variance and that the CT items also represent CT test specific variance which although overlapping on the

total test score level (r = .62) and somewhat on the subtest level (low and moderate CR and CT subtest correlations, Follman, Lowe, and Wiley, 2) apparently overlaps very little on the individual item level.

Rotation in the 102 x 102 analysis indicated substantially the same factor structure for CR and CT as was indicated in the 50 x 50 CR analysis and 52 x 52 CT analysis independently. The fact that the relatively homogeneous CR structure and the heterogeneous CT structure held up in the combined analysis is additional evidence of the disparate nature of CR and CT variance.

It is therefore concluded that critical reading and critical thinking do not strongly relate. This conclusion is veridical in so far as CR represents critical reading and CT represents critical thinking.

In order to infer the definition of critical reading, items loading .30 or above on the rotated large group factor in the 50 x 50 CR analysis were examined for apparent commonality. The large group factor was composed essentially of items 1, 4, 8, 9, 12, 13, 14, 16, 17, 20, 21, 28, 29, 32, 36, 37, 39, 40, 41, 44. These items represented all five CR subtests approximately proportionally. The location of items in different subtest categorizations implies that the different subtests measure different abilities and correspondingly that their respective items do also. It is suggested that the subtests while purportedly representing different skills, actually do not. Additional evidence for this interpretation is the moderate and high correlations between the CR subtests (Martin, 4).

In any case it appears that the large CR group factor represents judgments about statements of similar content in true false and multiple choice objective test form. In the true false format subjects related statements for accuracy to a passage and in the multiple choice format subjects related words or phrases for synonymic accuracy to statements. It is therefore concluded that the variance measured by CR represents underlying thinking activity of judging verbal material

in objective test form for accuracy of meaning synonymically to other verbal statement or passage material.

In order to infer the definition of critical thinking, items loading .30 or above in the 52 x 52 analysis were examined for commonality. CT consists of a heterogeneous collection of items representing a considerable number of different thinking activities rather than one general thinking activity or a few thinking activities. Since nearly all the rotated factors were small group factors each of few items from at least two subtests it is difficult to infer accurately the definition of the thinking activities they purportedly represent. Twenty-one items loaded on the largest CT group factor indicating some common variance but rotation split off 13 of these items onto other factors, apparently because in addition to the common variance they individually had enough additional idiosyncratic variance to become separate from the common group factor. Critical thinking as inferred from the different separate factors described above is seen as a composite of skills, particularly judgments of how statements relate to conclusions, interpretation of verbal statements, recognition of assumptions, hypothesis verification.

As discussed above critical reading and critical thinking are seen as having small overlap in terms of measuring the same underlying variable. The relatively small common variance may in fact reflect other commonalities such as the medium of language used by both, or similarity in content across tests.

Finally, it should be noted that the numbers of items (variables) exceeded the numbers of Ss in the factor analyses so the results of the factor analyses can only be viewed as tentative.

Conclusions

1. Both CR and CT are psychometrically sound instruments.

2. CR represents a relatively homogeneous underlying variable.

3. CT represents a number of different underlying variables.

4. Critical reading was inferred to be thinking activity involving judging verbal material in true false and multiple choice form for accuracy of synonymic meaning to other verbal statement or passage material.

5. Critical thinking was inferred to be less clearly defined as a composite of thinking skills including judgments of how statements relate to conclusions, interpretation of verbal statements, recognition of assumptions.

6. Critical reading and critical thinking as represented by CR and CT, respectively, overlap only moderately.

7. Finally, the results of this study are viewed as tentative since the numbers of variables (items) exceeded the numbers of subjects in the factor analyses.

References

1. ACE, American Council on Education, Test of Critical Thinking Form G, 1951.

2. Follman, J., Lowe, A. J., and Wiley, R. Correlational and factor analysis of critical reading and critical thinking - twelfth grade. Unpublished study, University of South Florida, 1969.

3. Follman, J., Wiley, R., Lowe, A. J., and Miller, W. Canonical and partial correlation of critical reading - critical thinking test scores - twelfth grade. Unpublished study, University of South Florida, 1969.

4. Martin, M. E. The Construction of a Diagnostic Reading Test for Senior High School Students and College Freshmen. Bureau of Publications, Teachers College, Columbia University, 1939.